DOCUMENT RESUME

ED 174 964 CS 004 958

TITLE Testing and Evaluation in Reading and Communication Skills: Abstracts of Doctoral Dissertations Published in "Dissertation Abstracts International," October 1978 through June 1979 (Vol. 39 Nos. 4 through 12).

INSTITUTION ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill.

PUB DATE 79

NOTE 20p.

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Cloze Procedure; *Composition (Literary); Doctoral Theses; *Educational Research; Elementary Secondary Education; Higher Education; Reading Ability; *Reading Achievement; *Reading Research; Reading Skills; Spelling; Standardized Tests; *Student Evaluation; *Testing; Test Reliability; Tests; *Test Validity; Test Wiseness; Writing Skills

ABSTRACT

This collection of abstracts is part of a continuing series providing information on recent doctoral dissertations. The 30 titles deal with a variety of topics, including the following: use of short and long passages in informal reading inventories, components in decoding spelling patterns, the psycholinguistic analysis of oral reading miscues and semantic acceptability, assessing reader ability to interpret internal punctuation, test wiseness and performance on standardized reading tests, the validities of selected tests, cloze procedure as a measure of reading/writing performance, oral language proficiency as related to reading readiness, techniques in analyzing a remedial reading program, word boundaries and reading achievement, the effectiveness of cloze tests and informal reading inventories, development of a communicative abilities test for young children, standardization and validation of the Horst Reversals Test, applying generalizability theory to writing assessment, predicting college student reading achievement, developing a reading attitude instrument, identifying and testing essential literal reading skills, the effects of reading test administration and score interpretation on student reading attitudes, the difficulty of comprehension questions in fourth grade reading tests, the assessment of high school student expository writing, measuring writing quality by T-units, assessing freshman composition, and readability formulas applied to college textbocks. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED174964

Testing and Evaluation in Reading and Communication Skills:

Abstracts of Doctoral Dissertations Published in Dissertation Abstracts International, October 1978 through June 1979 (Vol. 39 Nos. 4 through 12)

Compiled by the staff of the

ERIC Clearinghouse on Reading and Communication Skills

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
University Microfilms
International
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 004 952

The dissertation titles contained here are published with permission of the University Microfilms International, publishers of Dissertation Abstracts International (copyright © 1978, 1979 by University Microfilms International), and may not be reproduced without their prior permission.

This bibliography has been compiled as part of a continuing series designed to make information on relevant dissertations available to users of the ERIC system. Monthly issues of Dissertation Abstracts International are reviewed in order to compile abstracts of dissertations on related topics, which thus become accessible in searches of the ERIC data base. Ordering information for the dissertations themselves is included at the end of the bibliography.

Abstracts of the following dissertations are included in this collection:

Bowden, Nancy Elizabeth Butler
A STUDY OF FOURTH GRADE STUDENTS' READING COMPREHENSION MEASURES IN SHORT AND LONG PASSAGES OF AN INFORMAL READING INVENTORY

Campbell, John Olin, III
COMPONENT PROCESSES IN CHILDREN'S DECODING OF SPELLING PATTERNS

Carey, Robert Francis
A PSYCHOLINGUISTIC ANALYSIS OF THE EFFECTS OF SEMANTIC ACCEPTABILITY OF ORAL READING MISCUES ON READING COMPREHENSION

Carr, Mary Noreena
AN INSTRUMENT TO ASSESS READERS' ABILITY TO INTERPRET INTERNAL PUNCTUATION

Chang, Te Jung
TEST-WISENESS AND PASSAGE-DEPENDENCY IN STANDARDIZED READING COMPREHENSION TEST ITEMS

Derby, Thomas Lee
THE EFFECTS OF INSTRUCTION IN SELECTED ASPECTS OF TEST-WISENESS ON THE ADMINISTRATION OF STANDARDIZED READING TEST ITEMS IN THE UPPER ELEMENTARY SCHOOL

Duran, Elva
THE VALIDITY OF THE THREE DIAGNOSTIC READING TESTS

Erdman Lipner, Barbara
AN INVESTIGATION OF THE NATURE AND POSITION OF DELETIONS ON THE PERFORMANCE OF HIGH SCHOOL REMEDIAL READERS ON A CLOZE PROCEDURE

Euster, Sandra Duni
UTILIZATION OF THE CLOZE PROCEDURE AS A MEASURE OF WRITING SKILL OF COLLEGE STUDENTS

Gray, Richard Allen
THE RELATIONSHIP OF ORAL LANGUAGE PROFICIENCY IN FIVE AND SIX YEAR OLD PRESCHOOLERS TO READINESS FOR SCHOOL SUCCESS

Harvey, Leah Mary Stanton
AN ANALYSIS OF THE EFFECTIVENESS OF LEAST SQUARES, RIDGE AND ROBUST REGRESSION TECHNIQUES IN ANALYZING DATA FROM A REMEDIAL READING PROGRAM

Hecker, Nelly M. J.
AN INVESTIGATION OF WORD BOUNDARIES AS RELATED TO READING SKILL PERFORMANCE

Homan, Susan Lubet
DETERMINING INSTRUCTIONAL READING LEVEL: AN INVESTIGATION OF THE RELATIONSHIP BETWEEN STANDARD CLOZE TESTS, MULTIPLE CHOICE CLOZE TESTS, AND THE INFORMAL READING INVENTORY

Johnston, Elizabeth Booth
DEVELOPMENT OF THE COMMUNICATIVE ABILITIES TEST FOR YOUNG CHILDREN

Kaufman, Nadeen Laurie
REVERSALS AND READING: STANDARDIZATION AND VALIDATION OF THE HORST REVERSALS TEST FOR BLACK AND WHITE FIRST GRADE CHILDREN

Lagana, Janet Lee
THE EFFECTS OF INCENTIVE MOTIVATION AND TEST-WISENESS COACHING ON THE STANDARDIZED READING TEST SCORES OF THIRD-GRADE STUDENTS

Lindquist, Geraldine Taylor
THE VALIDITY OF THE DENVER DEVELOPMENTAL SCREENING TEST FOR PREDICTING READING ACHIEVEMENT OF PRIMARY GRADE STUDENTS

Llabre, Maria Magdalena
AN APPLICATION OF GENERALIZABILITY THEORY TO THE ASSESSMENT OF WRITING ABILITY

Magee, Patricia Ann
A STUDY OF THE VALIDITY OF FIVE SHORT FORMS OF THE TEST OF LANGUAGE DEVELOPMENT

Martin, Deanna Coleman
PREDICTING READING ACHIEVEMENT IN COLLEGE STUDENTS

Nolan, Marianne Casey
AN INVESTIGATION OF THE CLOZE PROCEDURE AS A MEANS TO FACILITATE GROUPING FOR READING INSTRUCTION

Pope, Jean Grove
A FIVE YEAR FOLLOW-UP STUDY EXAMINING THE VALIDITY OF A READING READINESS DIAGNOSTIC SCREENING BATTERY

Rhody, Regina Tullock
THE DEVELOPMENT OF A READING ATTITUDE INSTRUMENT FOR GRADES SEVEN THROUGH TWELVE

Roblyer, Margaret Dale
IDENTIFICATION AND TESTING OF ESSENTIAL SKILLS IN LITERAL READING COMPREHENSION

Shannon, Albert Joseph
SOME EFFECTS OF METHODS OF STANDARDIZED READING-ACHIEVEMENT TEST ADMINISTRATION AND SCORE INTERPRETATION ON SENIOR-HIGH STUDENTS' ATTITUDE TOWARD READING

Smith, Elizabeth Jane Wooten
DIFFICULTY OF COMPREHENSION QUESTIONS IN NORM- AND CRITERION-REFERENCED TESTS ADMINISTERED TO FOURTH GRADE PUPILS

Smith, Laura Spooner
AN INVESTIGATION OF THREE DOMAIN REFERENCED STRATEGIES TO ASSESS SECONDARY STUDENTS' EXPOSITORY WRITING

Stokes, Patricia Welch
THE QUALITY OF STUDENT COMPOSITION AS PREDICTED BY AVERAGE NUMBER OF WORDS PER T-UNIT AND ORGANIZATION SKILLS

Wallace, Rose Ann
ISSUES OF VALIDITY AND RELIABILITY IN THE TESTING OF FRESHMAN COMPOSITION

Wood, Westa Winn
A COMPARATIVE STUDY OF READABILITY FORMULAS APPLIED TO COLLEGE TEXTBOOKS

A STUDY OF FOURTH GRADE STUDENTS' READING COMPREHENSION MEASURES IN SHORT AND LONG PASSAGES OF AN INFORMAL READING INVENTORY

Order No. 7818219

BOWDEN, Nancy Elizabeth Butler, Ed.D. University of Houston, 1978. 179pp.

The current literature in reading reflects considerable controversy surrounding the informal reading inventory. Conflicting theories of the reading process have caused disagreements about recommended informal reading inventory procedures and evaluation criteria.

The purposes of this study were to:

1. Test the use of a 70% comprehension score in short passages as the basis of determining the oral and silent instructional levels.
2. Investigate differences in comprehension performance of readers of varying proficiency in short oral and silent reading passages.
3. Compare results of oral and silent reading diagnostic methods with comprehension performance in long passages.
4. Compare students' comprehension performance in short and long passages of equivalent difficulty.

Six hypotheses were proposed:

H_1 Students classified as "above grade level" in the short silent passages will have higher silent instructional levels than oral instructional levels.

H_2 Students classified as "on grade level" in the short silent passages will have the same silent and oral instructional levels.

H_3 Students classified as "below grade level" in the short silent passages will have higher oral instructional levels than silent instructional levels.

H_4 Students whose oral instructional levels are higher than their silent instructional levels will exhibit a higher level of comprehension performance in the long passages at the silent instructional levels than they will exhibit in the long passages at the oral instructional levels.

H_5 Students whose silent instructional levels are higher than their oral instructional levels will exhibit a higher level of comprehension performance in the long passages at the oral instructional levels when these passages are read silently.

H_6 The assignment of reading levels as determined by short silent passages will result in the correct placement of students in long silent passages at the same levels of difficulty.

A repeated measures design with stratified random sampling was used with a sample of 132 fourth graders. Groups were selected by the application of specific criteria for each analysis. The hypotheses were tested with a series of six t tests for paired measures at the .05 level of confidence.

All students were administered the Standard Reading Inventory (SRI), Form A, to determine their comparative proficiency as "above," "on," or "below grade level." All students were retested with the SRI, Form B, to determine their oral and silent instructional levels and their classifications as having higher oral or silent instructional levels. The students then read long passages drawn from three basal series at their oral and silent instructional levels.

Results of the analyses indicated that there were no significant differences in the oral and silent instructional levels of "above" and "on grade level" students, but that the "below grade level" students had significantly higher oral levels. The higher oral group had significantly higher scores in the long passages at the silent instructional levels, but were at frustration levels in the long passages at the oral instructional levels. The higher silent group had no significant differences in their long passage scores which were both at frustration levels. Significant decreases were found in comprehension scores in long passages at the silent instructional levels. Additional analyses indicated that all groups' comprehension scores in long passages showed significant decreases. When a 70% comprehension score in a short oral passage was used to determine placement in long passages, all groups had frustration level scores except for the "above grade level" group. Patterns of performance in the comparative proficiency groups suggested the possibility of differential diagnosis. The strong tendency of the sample to exhibit higher levels in oral reading suggested the need to reconsider reading demands at intermediate levels.

Recommendations were:

1. That the assumption of a shift towards silent reading proficiency at the fourth grade level be examined.
2. That experimental studies of strategies for teaching silent reading skills to "below grade level" readers be attempted.
3. That a 70% comprehension standard not be used as a basis for placement and that more stringent interpretation guidelines be followed.
4. That further study of the effects of comparative proficiency on reading performance be undertaken.
5. That investigations of placement tests other than IRI's be undertaken to compare test results with long passage performance.
6. That the reasons for the increased difficulty of long passages be explored.

COMPONENT PROCESSES IN CHILDREN'S DECODING OF SPELLING PATTERNS

Order No. 7822483

CAMPBELL, John Olin, III, Ph.D. Stanford University, 1978. 88pp.

This study investigated methods for analyzing reading into component processes by diagnostic tests that yield reliable and easily interpretable results. The research built upon the work of Calfee in developing "on-off" tests of components in which students tend to make either few or many errors.

A hierarchical on-off model of decoding performance was proposed in which the ability of a student to pronounce a given spelling pattern in the context of a word or pseudoword (decoding) was analyzed into a number of component processes. The processes were presumed to be hierarchically related and to be either learned (on) or unlearned (off). They were hierarchical in that if one or more processes were prerequisite to performance on a criterion test, then failure on a prerequisite would imply failure on the criterion test. Variations from all correct or none correct were assumed to be measurement errors. A double binomial model was hypothesized to characterize the total distribution of students in the learned and unlearned groups.

Spelling patterns were selected to span a range of difficulty, from consonant-vowel-consonant patterns like CAT, to blends (ST and SK), consonant digraphs (CH, SH and TH), vowel digraphs (EE, EA and OA) and the vowel-final e marker (e.g., MITE). Words and pseudowords were synthesized from the patterns plus consonants b, d, m, p and t.

Four tests were constructed, each designed to measure a different component:

a. Decoding (e.g., "Pronounce MEAT")

b. Pronunciation of the Isolated Pattern (e.g., "Pronounce EA")

c. Sound Comparison (e.g., "Do you hear /EAT/ in /MEAT/?")

d. Spelling Pattern Recognition (e.g., "Point to the letters which say /EA/ in /MEAT/.")

The design for each test consisted of two-between subject variables of grade and sex and the within-subject variables of spelling pattern, position within the word (consonant patterns only), wordness (word vs. pseudoword) and alternate forms.

Twenty-four students (12 male and 12 female) in the first and second grades were selected for testing.

The on-off aspect of the model was confirmed by visual inspection of frequency distributions for Decoding scores. Most patterns were strongly bimodal. Very easy patterns were skewed, with most students achieving a high score. Very difficult patterns were skewed in the opposite direction. A double binomial curve fit the distributions closely and yielded smaller Chi Squares than a single binomial.

The hierarchical aspect of the model was supported, in that very few students could decode who could not perform the component tasks, as evidenced in visual inspection of the scatterplots. A second analysis, which compared the product of components to their sum, did not permit clear support for the hierarchical aspect of the model because the product and the

sum were very highly correlated and did not make differential predictions.

The study provided direct evidence that on-off tests can make sharp diagnostic distinctions between students on decoding skills. Standard reading tests only indicate overall performance because they combine and therefore lose diagnostic information. Tests like those developed here can be used to evaluate components of reading performance for individual students and to pinpoint areas for training.

A PSYCHOLINGUISTIC ANALYSIS OF THE EFFECTS OF SEMANTIC ACCEPTABILITY OF ORAL READING MISCUES ON READING COMPREHENSION Order No. 7911342

CAREY, Robert Francis, Ph.D. The University of Connecticut, 1978. 194pp.

Purpose

This study explores relationships between patterns of oral reading performance and performance on indicators of reading comprehension. The study seeks to determine the efficacy of using combinatorial descriptors of reading behavior, each consisting of characteristics drawn from the Goodman taxonomy of oral reading miscues, in predicting student performance on three comprehension measures.

The characteristics employed are the degree of semantic acceptability of the miscue and the type of overt correction behavior exhibited. Five degrees of semantic acceptability and four types of correction behavior are considered. Reading comprehension is inferred from performance on a standardized reading comprehension test, a post oral-reading cloze test, and a comprehending score. The comprehending score is a process measure indicating the proportion of miscues suggesting a successful search for meaning on the part of the reader.

Procedures

Miscue analyses were performed on audiotape recordings of the oral readings of 100 students in grades 6 through 9 of a suburban Rhode Island school district. A stratified random sampling procedure was used to select the subjects for the study. Each miscue was accorded to one of twenty categories on the basis of degree of semantic acceptability and type of correction behavior. Also recorded were each student's post oral reading cloze test score, standardized comprehension test score, and comprehending score.

Multiple linear regression analysis provided the framework for examination of the data. The relative frequencies of the twenty miscue types served as independent variables. Student scores on the three comprehension indicators served as dependent variables. The analysis provided for the attribution of specific proportions of criterion variance to each of the miscue categories. Order of entry of the independent variables was dictated by a combination of hierarchical and step-wise inclusion techniques. The data analysis focused on the percent of variance shared by independent and dependent variables. Common variance was inferred from the degree of change in coefficients of determination of adjacent variables. Scatterplots were generated and a variety of function transformations were employed in analyses for curvilinear trends.

Conclusions

The data suggest that oral reading miscues, when considered on the basis of degree of semantic acceptability and correction behavior, maintain a persistent and significant relationship with reading comprehension performance. The relative frequencies of several miscue types in oral reading performance may be inferred as indicative of the adequacy of reading comprehension. The data further suggest that the miscue types studied can be hierarchically arranged according to a theoretical altercue continuum; i.e., some types positively predict comprehension, others negatively predict comprehension, and several types display entropic characteristics.

The findings suggest support for the construct validity of the comprehending score and tend to support earlier research which posited the importance of semantically acceptable miscues as indicators of reading proficiency. The findings also support the integral function of semantic monitoring in psycholinguistic paradigms.

The data suggest the correlational validity of the post oral-reading cloze test as an index of comprehension performance. Also, a persistent relationship between student miscue performance and cloze performance is implicit in the findings.

The data further suggest that comprehension is not a uniform factor reflected in the same manner or to the same degree among the measures. However, a measurement factor common to the three comprehension indicators may be inferred from the data. These findings suggest the feasibility of a comprehension assessment device, appropriate for classroom or clinic, which employs the semantic acceptability and correction behavior criteria.

AN INSTRUMENT TO ASSESS READERS' ABILITY TO INTERPRET INTERNAL PUNCTUATION Order No. 7823171

CARR, Mary Noreena, Ed.D. Arizona State University, 1978. 79pp.

The relationship of internal meaning-bearing punctuation with silent reading comprehension was investigated. A measure to test this interrelationship was constructed, and designated as the Punc Test. The test contained a passage which had 20 pairs of matched sentences imbedded in it. One sentence of each pair contained internal meaning-bearing punctuation, while the matched sentence required no internal punctuation. Seventh-grade readability level was established for the passage through application of the Dale-Chall and Fry formulas. Questions directed to the literal level of the 40 sentences were constructed. The test was piloted, revised, and later administered to 255 seventh graders. A Spearman-Brown split-half reliability of .73 was obtained. Construct validity was established through analysis of data rendered by the two subtests. This confirmed the theory that students who are unable to interpret internal meaning-bearing punctuation are impeded in the comprehension of a written passage which contains such punctuation. Concurrent validity for general comprehension was obtained through correlation with the SRA Survey Assessment reading comprehension test. It was concluded that the Punc Test could determine a reader's ability to interpret internal punctuation of a written passage.

TEST-WISENESS AND PASSAGE-DEPENDENCY IN STANDARDIZED READING COMPREHENSION TEST ITEMS Order No. 7908015

CHANG, Te Jung, Ph.D. Southern Illinois University at Carbondale, 1978. 147pp. Major Professor: Dr. Ernest L. Lewis

Standardized reading comprehension tests, ordinarily, are thought to be passage-dependent. Students must read through the reading passages in order to correctly answer the test items. However, researchers have found that many standardized reading comprehension test items can be correctly answered by students without their necessarily reading the related passages. They have attributed students' capabilities of doing this to lack of passage-dependency of the test. While their assertions were not groundless, they seemed to have ignored another possibly important factor--the students' test-taking skills, i.e., test-wiseness which may unexpectedly benefit students to answer the reading comprehension test items correctly.

The purposes of this study were to investigate (1) whether it was test-wiseness or passage-dependency that caused the subjects to answer test items in a standardized reading comprehension test correctly when its related reading passages

were not provided to them; (2) whether test-wiseness affected the subjects' reading comprehension upon repeated measures; and (3) whether the stability of the subjects' reading comprehension scores was related to their test-wiseness abilities.

One hundred and fourteen undergraduate students at Southern Illinois University in the Spring Semester, 1978 served as the subjects of this study. They were randomly assigned to three groups: the test-wiseness group (T-W), the passage-independent group (P-I), and the control group, with 38 subjects each. The T-W group was given a specific test-wiseness training so that the subjects in this group would be more sophisticated in test-taking skills than the subjects in the other two groups which were not given test-wiseness training.

Two criterion measures employed in this study were the Standard Educational Intelligence Test (SEIT) and the Nelson-Denny Reading Test--Comprehension, the revised Form A. Subjects of all three groups were given the SEIT first and then the Nelson-Denny test in pre- and post-test situations at three-week intervals. Both the T-W and the P-I groups were given the Nelson-Denny test with passages deleted. The control group was given the same test with passages intact.

A 3 x 2 factorial design with repeated measures on testing dimension was used to test the hypotheses related to the first two purposes of this study. The third purpose of this study was served by testing the hypotheses concerning the correlations of test-wiseness and reading comprehension for each group and between any two groups.

The results of this study indicate that while the T-W group scored significantly higher than the P-I and the control groups on the pre- and post-test-wiseness tests, the T-W group did not perform any better than the P-I group on the pre- and post-reading comprehension tests. The results confirm the findings of previous research that some standardized reading comprehension test items do seem to lack passage-dependency and can be correctly answered by students without necessarily reading through its related passages. Further, the mean pre- and post-reading comprehension test scores of these three groups did not manifest the definite patterns as was expected.

Whether the causes of lack of passage-dependency are due to test item flaws or to students' test-wiseness cannot be determined at this point. However, it seems reasonable to conclude that the P-I subjects who were able to score slightly higher than the T-W subjects might have greatly relied upon their natural test-wiseness abilities even though their performances on the test-wiseness test were lower than the T-W subjects. The T-W subjects might have excessively applied the test-wiseness principles to the reading comprehension test situations to the extent that the tendencies of utilizing their natural analytical abilities to solve the test questions were hindered.

Although the generalizability of the test-wiseness principles to external test situations seems to be questionable, the results of this study show that test-wiseness and passage-dependency seem to be inseparable. The results of this study also indicate that the subjects' test-wiseness abilities do relate to their reading comprehension test performances. When the subjects were given the reading comprehension test without passages, they tended to rely more upon their test-wiseness abilities than those subjects who were given the same test with passages.

THE EFFECTS OF INSTRUCTION IN SELECTED ASPECTS OF TEST-WISENESS ON THE ADMINISTRATION OF STANDARDIZED READING TEST ITEMS IN THE UPPER ELEMENTARY SCHOOL Order No. 7911904

DERBY, Thomas Lee, Ed.D. University of Pennsylvania, 1978. 141pp. Chairman: J. Wesley Schneyer

This experimental investigation was designed to provide answers to several questions about one aspect of test-wiseness (TW), namely, the ability of elementary pupils to use extraneous clues in answering multiple-choice standardized reading-comprehension test items. More specifically, the study was designed to answer six questions that are related to fifth- and sixth-grade pupils' ability to answer multiple-choice items from the Stanford Achievement Test, Intermediate Levels I and II, Reading Section, with reading passages removed: (1) Can children with no instruction in the use of item clues scores higher than chance on passageless Stanford (R-P) items? (2) Can children who have received instruction in using three extraneous item clues (absurd options, clues in stems, and clues in other items) score higher on R-P items than similar uninstructed children? (3) Can male children score higher than female children on R-P items? (4) Can children who have been pretested with a form of the R-P instrument score higher on a parallel form of the R-P instrument, given as a posttest, than similar unpretested children? (5) How well can graduate students with some background in educational measurements use item clues to identify the R-P items that children can answer correctly? (6) How is children's general reading ability as measured on the Metropolitan Achievement Test related to their ability on the R-P items?

The overall plan of the investigation included a successful procedure consisting of two pilot studies for evaluating the R-P test and revising the clue-using TW instructional program prior to their inclusion in the Main Study. The pilot studies were also important in that some of their data were used in testing the research hypotheses. The Main Study research design allowed for measuring simultaneously the effects of instruction, pretesting, and gender, and it included a statistical procedure, analysis of covariance, to adjust for initial differences among subjects in reading ability.

Children who participated in the investigation were mostly from middle- and upper-middle-class suburban neighborhoods.

Pilot studies as well as the Main Study yielded important findings. The R-P instrument was found to be a moderately reliable ($r = .55$, $p < .01$) test of TW for Pilot Study I graduate students. Pilot Study II and Main Study children with no instruction in the use of item clues scored significantly ($p < .001$) higher on the R-P instrument than the score that would be expected on the basis of chance, i.e., blind guessing. In the Main Study, the main effects of TW instruction and gender had significant ($p < .001$) effects as measured by the analysis of covariance; however, since a significant ($p < .05$) instruction-by-gender interaction was also revealed by the analysis of covariance, the corresponding main effects were not considered. The effect of pretesting upon children's R-P scores was nonsignificant. Graduate students' correct R-P responses that were based upon item clues correlated significantly ($p < .005$) with children's correct R-P responses. Children's general reading ability as measured by raw scores for Total Reading on the Metropolitan Achievement Test were correlated significantly ($r = .64$, $p < .005$) with their R-P scores.

The findings indicated that for participating fifth and sixth-graders most of the R-P items were passage-independent, i.e., susceptible to being answered correctly without referring to their corresponding passages. It appeared that graduate students' ability to identify the passage-independent items on the basis of item clues could be helpful in future improvements of standardized tests. Though it appeared that both TW instruction and gender had had an effect on R-P scores, the significant instruction-by-gender interaction precluded acceptance of the hypotheses involving those variables and indicated that further research would be needed to explain why, as the interaction seemed to indicate, males profited more from the TW instruction than females.

THE VALIDITY OF THE THREE DIAGNOSTIC READING TESTS Order No. 7901070

DURAN, Elva, Ph.D. University of Oregon, 1978. 156pp. Adviser: Ruth Waugh

The purpose of this study was to validate three diagnostic phonic reading tests in order to determine which test best predicted what a child did know in phonics.

The design of the study consisted of taking 24 phonic elements which were found common among the three diagnostic reading tests. The reading tests used in the study were: The El Paso Phonics Survey (an individual measure testing the phonic elements by using nonsense words), the Botel Reading

Inventory (the measure has pupils write the correct element or word being tested) and the Eugene Scope and Sequence Test (which has pupils select the appropriate answer). The three tests were validated by using two criterion measures. Both criterion measures represented the 24 phonic elements under consideration. Criterion Measure I consisted of a reading passage while Criterion Measure II represented a word list.

All measures were administered during a three week period to 108 second graders in the Eugene School District.

The Pearson Product Moment Correlation was calculated. Results for the three tests and the three tests and the criterion measures are shown below:

TABLE I. Correlation Coefficients Among the Three Diagnostic Reading Tests

	El Paso	Botel	Eugene	C.M. I	C.M. II
El Paso	1.000	0.502	0.440	0.666	0.783
Botel	0.502	1.000	0.221	0.432	0.528
Eugene	0.440	0.221	1.000	0.309	0.376

CM I and CM II = criterion measures

In the analysis the Pearson Product Moment Correlation revealed the El Paso Phonics Survey to be the more highly correlated with the criterion measure than either the Botel or Eugene Scope and Sequence Tests.

Reliability coefficients for the Botel Reading Inventory and the El Paso Phonics Survey were shown to be moderately high. The coefficient alpha revealed (0.866) for the Botel Reading Inventory, (0.813) for the El Paso Phonics Survey and the lowest coefficient was reported for the Eugene Scope and Sequence Test at (0.572).

It is clear that the three tests are all measuring different skills as determined by the correlation coefficients among the tests and criterion measures. Finally, it is evident that the El Paso Phonics Survey, an individual assessment made of all nonsense words, is the best predictor of what children do know in phonics.

AN INVESTIGATION OF THE NATURE AND POSITION OF DELETIONS ON THE PERFORMANCE OF HIGH SCHOOL REMEDIAL READERS ON A CLOZE PROCEDURE

Order No. 7821529

ERDMAN LIPNER, Barbara, Ed.D. Hofstra University, 1978. 175pp.

The purpose of this investigation is to explore the effect of the deletion of content words and function words as well as the effect of the position of the deletion in the sentence on the successful completion of cloze passages (fourth grade readability level) by one hundred high school remedial readers who score between 5.0 and 6.9 on the Metropolitan Reading Test, Advanced-Form H.

To accomplish this purpose, the Dale-Chall Readability Formula was used to select four passages at the fourth grade reading level. Cloze tests were constructed by leaving the first and last sentences intact, randomly selecting the first word to be deleted from the second sentence, and then deleting every fifth word. Exact replacements which did not have more than two deviations from a correct phonic representation were accepted.

An analysis of variance was computed to determine if a significant difference existed among the percentage of correct student responses on the total cloze score, on the correct replacement of content words within the passage, on the correct replacement of function words within the passage, and on a multiple choice comprehension test. All these factors were found to be significantly different, so the Tukey A Test of Multiple Comparisons was computed to compare (1) content words and function words, (2) function words and cloze score, (3) content words and cloze score, (4) cloze scores and comprehension, (5) function words and comprehension, and (6) content words and comprehension. The Tukey A calculated the comparison of content words and cloze scores to the .05 level of significance for two of the passages. All other comparisons were significantly different at the .01 level of significance.

A Chi Square analysis was computed to compare the percentage of function words and the percentage of content words in the 40 percent easiest deletions and the 40 percent most difficult deletions. There was a significant difference, with function words occurring significantly more frequently in the easy deletions and content words appearing significantly more often in the difficult deletions.

Using a Chi Square analysis, a significant difference at the .05 level was found to exist between the percentage of the 40 percent easiest deletions which occur in the first third of the sentence and the percentage of the 40 percent most difficult deletions which occur in the first third of a sentence. The same analysis was computed for the last third of a sentence, and significance was at the .01 level.

A Spearman Rank Order Correlation was calculated to test for a relationship between the percentage of content words deleted and the difficulty of the cloze passage, and to test for a relationship between the percentage of function words deleted and the difficulty of the cloze passage. In neither case was a significant difference computed.

Finally, a Chi Square Analysis indicated no significant difference between the percentage of content words and function words deleted in passages of increasing difficulty.

Among the conclusions reached were the following: (1) Successful completion of cloze blanks, correct replacement of function words, and correct responses on a multiple choice comprehension test are all independent variables measuring different elements of reading; (2) Content words are more difficult to replace correctly than are function words; (3) Content words occur more frequently in the last third of a sentence and are more difficult to replace; (4) Deletions which occur in the first third of a sentence are easier to replace; (5) There is no difference in the percentage of content words and function words deleted by a random every fifth cloze procedure at grade levels four, five, seven, and nine.

UTILIZATION OF THE CLOZE PROCEDURE AS A MEASURE OF WRITING SKILL OF COLLEGE STUDENTS

Order No. 7907601

EUSTER, Sandra Duni, Ph.D. University of South Carolina, 1978. 102pp. Major Professor: Dr. Paul C. Berg

The purpose of this study was to seek answers to the following research questions:

I. (A) What is the relationship between cloze test scores and writing sample outcomes?

(B) To what extent are cloze test scores and writing samples related when controlling for achievement, race, sex, major, and year in school?

II. (A) What is the relationship between cloze test scores and reading skills as measured by:

(1) The Nelson Denny Reading Vocabulary Test,
(2) The Nelson Denny Reading Comprehension Test
(3) The entire Nelson Denny Test?

(B) To what extent are cloze test scores and reading skills related when controlling for achievement, race, sex, major, and year in school?

III. (A) What is the relationship between writing sample outcomes and reading skills as measured by:

(1) The Nelson Denny Reading Vocabulary Test,
(2) The Nelson Denny Reading Comprehension Test,
(3) The entire Nelson Denny Reading Test?

(B) To what extent are writing sample outcomes and reading skills related when controlling for achievement, race, sex, major, and year in school?

The approach utilized in this study was to select all the students from four freshmen English classes at The College of General Studies, University of South Carolina and administer the following tests: (a) The Nelson Denny Reading Test, (b) A cloze test, (c) An essay test. A total of 80 students were involved in the study.

Data were analyzed at the computer center, University of South Carolina employing the SPSS computer program (Statistical Package for the Social Sciences). A Pearson Product Moment Correlation was utilized to determine the relationship between the cloze test scores and writing, between the cloze/test scores and reading skills, and between the writing sample outcomes and reading skills. A partial correlation was employed between the measures listed above when a third variable was added.

The analysis of the data revealed the following results.

Question I

There was a significant positive relationship between cloze test scores and writing sample outcomes. Additionally, achievement had a significant positive relationship, however, the correlation was rather small. Correlations were significant for whites, but not for black students. However, this could have been due to sample size. Criminal Justice, Commercial Education, and Undecided majors had significant correlations. Correlations for freshmen were significant, but were not for sophomores. Sample size may have played a predominant part here also.

Question II

Another major finding indicated there was a strong positive relationship between the cloze test and reading skills as measured by The Nelson Denny Reading Test. When controlling for achievement it was found that there was a significant positive relationship. Females had correlations significant for vocabulary, comprehension, and the total test, while males had significant correlations for vocabulary and the total test.

All correlations were significant for whites. The only significant correlations for blacks involved the one for the total test. Again sample size may have influenced this.

All correlations for freshmen were significant. None were significant for sophomores. Commercial Education and Child Development majors had significant correlations for all parts of The Nelson Denny Reading Test.

Question III

There was a significant positive relationship between writing sample outcomes and reading skills. Additionally, the correlation for achievement was significant for both vocabulary and total test.

Females had significant correlations, while males did not have any correlations which were significant. The correlations for white students were all significant, while they were not for black students. All correlations were significant for freshmen, while only the correlations for comprehension were significant for sophomores.

All correlations were significant for Undecided majors, while only vocabulary scores were significant for Criminal Justice majors.

The findings of this study indicate that writing and reading are quite global in nature, and one has a great effect on the other. Additionally, the cloze test can now be used with a battery of tests to assess writing and reading skill.

THE RELATIONSHIP OF ORAL LANGUAGE PROFICIENCY IN FIVE AND SIX YEAR OLD PRESCHOOLERS TO READINESS FOR SCHOOL SUCCESS Order No. 7900570

GRAY, Richard Allen, Ph.D. The University of Texas at Austin, 1978. 118pp. Supervisor: Stephen C. Larsen

Receptive and expressive oral language play a significant role in programming and in identification of "severe discrepancy" at the early childhood and kindergarten levels. An important factor underlying this emphasis is the belief that a significant relationship exists between oral language proficiency and later school success--especially in the academic areas of reading, writing, and math. Although a body of theoretical and empirical literature can be cited to support this belief, other studies raise doubts, and the relationship remains inconclusive. The purpose of this study was to examine further the nature of the relationship of measures of receptive and expressive features of oral language to readiness and predicted school success.

The Test of Language Development (TOLD), a standardized measure of oral language proficiency, and the Basic School Skills Inventory (BSSI), a standardized measure of readiness, were administered to 70 five- and six-year-olds enrolled in five preschools in Austin, Texas. Tests were administered in three federally funded Child, Incorporated preschool centers and two private preschools to all students falling in the appropriate age range.

Analysis was conducted using correlation, regression and discriminant analysis. Results indicate that there is a moderate relationship between oral language proficiency and potential for academic achievement. In particular, receptive language functions appear to be related to scores on school readiness measures. Both receptive and expressive language abilities discriminated between children with high and low measured readiness.

This research would seem to support the hypothesis that oral language proficiency and academic achievement are related. Implications are that "severe discrepancy" in measured oral language skill is one indicator, at the preschool level, of potential learning problems in the school setting, and that assessment of oral language has meaning beyond the immediate, practical value of adequate communication skills. This has significance in the search for an adequate battery of tests for early identification of learning problems but should not be extended to the assumption that training in oral language will necessarily improve performance in other academic skills.

AN ANALYSIS OF THE EFFECTIVENESS OF LEAST SQUARES, RIDGE AND ROBUST REGRESSION TECHNIQUES IN ANALYZING DATA FROM A REMEDIAL READING PROGRAM Order No. 7912016

HARVEY, Leah Mary Stanton, Ph.D. University of Minnesota, 1978. 238pp. Chairman: Raymond O. Collier, Jr.

Linear regression is a statistical technique which is frequently used in educational research. The usual linear regression technique is called least squares regression. Two of the assumptions usually made when drawing inferences from least squares regression results are (a) that the predictor variables are not highly correlated and (b) that the prediction errors are approximately normally distributed. When these assumptions are not met, other linear regression techniques can be more effective than the least squares technique.

Ridge regression is a linear regression technique which has been used to analyze data for which the predictor variables are highly correlated. Robust regression techniques have been developed to analyze non-normal data. The purpose of this study was to examine the usefulness and efficiency of least squares, robust and ridge regression techniques in the analysis of data generated by a reading program for students with learning disabilities.

Two sets of variables were used in the analyses. In one, the dependent variable, the progress made by a student during the year was predicted by variables which were based upon

early performance in the reading program. The second set of variables was based on performance throughout the year; these variables were used to predict year end test scores on a standardized reading test.

Two cross-validation procedures were used to compare the three regression techniques. In the first, one observation at a time was deleted from each analysis and predicted using the resulting regression coefficients. In the second procedure, half of the observations were randomly deleted and then predicted using the coefficients resulting from the analysis of the remaining observations.

The effectiveness of each technique was measured by (a) the correlation coefficients for the deleted observations and their predicted values, (b) the sums of the squared differences between those values and (c) the sums of the absoluted differences between them.

The results of the study indicate that students' yearly progress could be predicted by early performance in the sequence. Three variables explained 74% of the variance in the dependent variable. The best predictor of posttest scores was pretest scores--this predictor alone accounted for 74% of the variance in the dependent variable.

The robust techniques were more effective in the analysis of the first data set than were the least squares techniques. It was comparable to the least squares technique in the second analyses where the data was more normal. The ridge technique used was comparable to, or slightly better than, the least squares techniques in the analyses of both data sets which included highly correlated predictors.

AN INVESTIGATION OF WORD BOUNDARIES AS RELATED TO READING SKILL PERFORMANCE Order No. 7914031

HECKER, Nelly M. J., Ph.D. University of Georgia, 1978. 157pp. Supervisor: George E. Mason

One purpose of this investigation was to examine how level of reading acquisition related to performance on word boundary tasks written at four differing readability levels. A second purpose of this study was to explore the validity of word boundary tasks as measures of reading ability by comparing student performance on a word boundary measure to performance on a cloze procedure task and on a test of word recognition.

The subjects for this study were 80 third grade students, 75 fifth grade students, and 98 seventh grade students selected from the Clarke County (Georgia) schools. The students were stratified by reading level and placed in "high," "average," and "low" treatment groups.

A split-plot factorial ANOVA design incorporating two independent variables and one repeated measure tested the following hypotheses:

1. Scores on the word boundary task will not differ as a main effect of reading acquisition level for: (a) third grade students, (b) fifth grade students, and (c) seventh grade students.
2. Scores on the word boundary task will not differ as a main effect of the readability level of the task among: (a) third grade students, (b) fifth grade students, and (c) seventh grade students.
3. There will be no significant interaction between word boundary readability levels and reading acquisition levels among: (a) third grade students, (b) fifth grade students, and (c) seventh grade students.

All hypotheses were rejected ($p < .001$).

The validity of using the word boundary task as a measure of reading ability was determined with a multitrait-multimethod matrix of intercorrelations between the informal measures.

The data revealed that level of reading acquisition, level of passage readability, and grade placement of students affected performance on the word boundary task. The well-defined pattern of greater means for "high" groups and smaller means for "low" groups, the fact that means for performance on the word boundary task increased with grade level, and the fact that word boundary scores declined as passage readability increased, support the conclusion that the word boundary tasks were a measure of reading performance for these students.

The high degree of correlation between the word boundary task and the cloze procedure, and between the word boundary task and the test of word recognition indicates that the three methods measure some of the same traits. Since the highest correlation values were found in levels of these measures that were closer to the grade placement of students, it was concluded that the more valid and reliable word boundary scores were those for passages at or near the reading level of the subjects.

DETERMINING INSTRUCTIONAL READING LEVEL: AN INVESTIGATION OF THE RELATIONSHIP BETWEEN STANDARD CLOZE TESTS, MULTIPLE CHOICE CLOZE TESTS, AND THE INFORMAL READING INVENTORY

Order No. 7913285

HOMAN, Susan Lubet, Ph.D. The University of Florida, 1978. 162pp. Chairman: William R. Powell

The relationship of a new form of cloze test, multiple choice cloze (MCC), to standard cloze and the Informal Reading Inventory (IRI) was explored in this study. The intent was to provide new information that would assist the classroom teacher in determining the instructional reading level of all students as easily and accurately as possible.

A secondary purpose of the study involved a comparison of MCC tests. The readability levels of the MCC passages were determined by traditional methods and by a new system of readability determination, Rasch calibration.

A third aspect of this study focused on placement decisions based on the two types of scoring criteria used with the IRI to determine instructional reading level.

Second, fourth, and sixth graders were participants in the study. Similar results at all three grade levels suggest generalizability of results for elementary and intermediate grade levels.

Correlations between standard cloze and MCC were low (ranging from $r = .27$ to $.80$), considering the same students were given the same stories in standard cloze and MCC forms. These results raise some question as to whether both forms are measuring the same type of reading comprehension.

High positive correlations were found between MCC passages with readability levels determined by traditional formula and MCC passages with readability levels based on Rasch calibration, indicating that these two methods of readability determination yield similar results.

A significant difference existed in placement of students by the Powell and Betts IRI criteria. These differences in placement indicate that classroom teachers should carefully choose the IRI scoring criteria they will use based on a conviction of accuracy of placement.

Three major implications for future research and practice were derived from this study. There is some evidence that MCC does not measure the same type of reading comprehension as measured by standard cloze or the IRI. The information on this issue is inconclusive and further study is indicated before more specific conclusions can be reached. The relationship between standard cloze and the IRI also appears tenuous. The scoring criteria used for the IRI can make very significant differences in terms of accurate placement.

DEVELOPMENT OF THE COMMUNICATIVE ABILITIES TEST FOR YOUNG CHILDREN Order No. 7904784

JOHNSTON, Elizabeth Booth, Ed.D. University of Cincinnati, 1978. 223pp.

Because language is basic to all other school learning, teachers and clinicians necessarily devote much time and effort to determining which children do and which children do not have language problems.

To provide a complete evaluation of communication, a test must enable the clinician to do several things:

1. To provide stimulus materials and format appropriate to the goals of the test,
2. To test children's spontaneous conversation in context,
3. To test generated language, not memorization or imitation,
4. To evaluate development in the basic areas of language,
5. To identify those children who have language problems, and
6. To discover norms for the children in the specified age range.

Investigation of the most widely used currently available commercial language tests did not reveal any which met all of the requirements stated above.

The Experimental Goal of this study was to construct a communication performance test for children ages 3 through 6 years, which would allow assessment of the development of communicative abilities in three of the four areas of communication: pragmatics, semantics, and syntax.

The test instrument was specifically constructed to meet the needs established for this research. It is composed of 3 Settings: A Story Setting, (during which the clinician makes observations of the child's use of syntactic and semantic structures), a Game Setting (during which the clinician observes the child's pragmatic ability to use language as a tool), and a Conversation Setting (during which the clinician makes observations of the child's use of discourse). The alternate story format was designed to generate spontaneous conversational language from the children.

A total of 136 children was used in the study: one group of 96 normally developing children, 24 at each of ages 3, 4, 5 and 6, and one group of 40 language impaired children divided among the same ages.

The normally developing population was tested first and the resulting data were used to determine the specific communicative content for each age level of the test, and to establish tentative 'norms' for this test.

An item analysis of each item was made and the items that were successful with at least 60 percent of the children who were expected to pass them were retained in the final form of the test. Items which were less than 60 percent successful with the children expected to produce them were abandoned. On the basis of this item analysis the test was modified and the resulting test was given to the 40 language impaired children to see if they would be correctly identified by this test instrument.

The test effectively discriminated between the normally developing children and the language impaired children. The test identified 39 of the 40 language impaired children in the appropriate range and the remaining one was in the borderline range. Of the 96 normally developing children, 2 were in the borderline range and two would have failed this test.

REVERSALS AND READING: STANDARDIZATION AND VALIDATION OF THE HORST REVERSALS TEST FOR BLACK AND WHITE FIRST GRADE CHILDREN Order No. 7908999

KAUFMAN, Nadeen Laurie, Ed.D. Columbia University Teachers College, 1978. 174pp. Sponsor: Professor Margaret Jo Shepherd

Reversal errors like b-d or was/saw confusions are commonly found both with beginning readers and with older children who suffer from severe reading disabilities. Orientation, a low salience cue to which young children gradually develop selective attention, is the prime distinctive feature of letters and letter orders (according to Gibson) which provides the relational information necessary to keep a child from responding to reversal and sequencing tendencies. For the older disabled reader, reversals have been shown to be a characteristic of some types of dyslexia. Tasks which evaluate a child's tendency to reverse have been able to predict future reading achievement. Despite the importance of the reversal phenomenon, there were no well standardized and validated tests of reversals for children. Hence, the main goal of this investigation was to standardize and validate the Horst Reversals Test, an instrument that showed promise in studies of de Hirsch's Predictive Index. A second goal of this study was to explore the differential predictive validity of the Horst test for blacks versus whites.

The sample comprised 401 first grade children (mean age = 6-4) in a Georgia school district, divided about equally by sex and race, and spanning a wide range of socioeconomic backgrounds. White children scored significantly higher than black children, so separate racial norms were developed ($N = 175$ for each group) in accordance with the latest Census data. The normative tables were stratified by sex, socioeconomic status (parental occupation), urban-rural residence, and prior schooling. The Horst test proved reliable for both racial groups, and correlated significantly, but moderately, with the Lee-Clark Reading Readiness Test, the Language and Mathematics subtests of the Tests of Basic Experiences, and the Draw-A-Design and Draw-A-Child subtests of the McCarthy Scales. Predictive validity coefficients for the Horst test (seven month interval), using the Metropolitan Achievement Tests and level in a reading series as the criteria, were in the .60s for the total group and the blacks, and in the .50s for the white children. A comparison of the predictive validity coefficients for blacks and whites yielded significant differences in favor of the blacks for two of the three reading related subtests of the Metropolitan Achievement Tests; no significant differences were obtained for the reading series criterion.

Sex differences were also explored. White girls scored significantly higher than white boys on the Horst test, but black boys and girls performed equally well. Predictive validity coefficients for boys and girls did not differ significantly for either blacks or whites.

Race differences in mean Horst test performance seemed largely attributable to discrepancies between blacks and whites on the variables of socioeconomic status and schooling prior to first grade. The moderate correlations between the Horst test and other ability tests was seen as indicative of both optimum commonality and uniqueness for each test evaluated. The predictive validity coefficients for the Horst test compared quite favorably to the results found in the literature for other reversals tasks. The key results of these analyses were that the Horst test was a very good predictor of reading achievement for both blacks and whites; seemed to be an unbiased measure based on the definition of test bias used in this investigation; and showed promise as a screening instrument for beginning first graders.

THE EFFECTS OF INCENTIVE MOTIVATION AND TEST-WISENESS COACHING ON THE STANDARDIZED READING TEST SCORES OF THIRD-GRADE STUDENTS

Order No 7901373

LAGANA, Janet Lee, Ed.D. University of California, Los Angeles, 1978. 113pp. Chairman: Professor John D. McNeil

This study was aimed at contributing to understanding the meaning of standardized reading achievement scores. Through much of the history of education, standardized testing has been accepted without question. Recently, however, much controversy, mixed emotions, and confusion have developed in our society as to the role of these tests in curriculum making and student placement.

The purpose of this study was three-fold. First, to determine if the standardized reading test scores of third-grade students are affected when an incentive is offered in an attempt to bring about maximum effort on the part of the student. Second, to determine whether or not teaching a unit on Testwiseness Tips would increase reading achievement test scores above the point of offering an incentive. Third, to determine if lower achieving third-grade students would profit more from additional motivation and instruction in testwiseness than would higher achieving students.

Two hundred eighty-four third-grade students were randomly assigned to one of three treatment groups. The first group, the Control Group, was administered the Vocabulary and Comprehension sections of the Comprehensive Tests of Basic Skills, Level 1, Form S, under normal conditions. The second group was told that they would receive an ice cream for trying to do their best on the same standardized reading test. The third group was given instruction on testwiseness principles, in a one-hour lesson prior to the testing, and were told on the day of the test that they would receive an ice cream if they used these principles while taking the test. Students within the 17 third-grade classrooms were ranked for achievement level on the total reading pretest score of the Comprehensive Tests of Basic Skills, which was administered in both October and May of the 1977-78 school year. All groups were administered the test at the same time, however not in the same room.

The results showed that incentive motivation, or the offering of an ice cream as a reward, had a statistically significant effect on the standardized reading tests scores of these third-grade students. In addition, the effect of incentive motivation and testwiseness training in combination was statistically significant. The effect of achievement level was significant at the .01 level. Additional findings were that boys did better with testwiseness training than did girls, but girls did better with an incentive motivation treatment than did boys.

Six recommendations were made concerning future research on testwiseness and incentive motivation in combination and ways in which the present study would be effectively replicated. Probably the most important statement that can be made as a result of this study is that it appears that standardized reading test scores are not revealing the achievement capabilities of students. Through testwiseness training and incentive motivation it has been possible to more fully reveal these capabilities.

THE VALIDITY OF THE DENVER DEVELOPMENTAL SCREENING TEST FOR PREDICTING READING ACHIEVEMENT OF PRIMARY GRADE STUDENTS Order No. 7821143

LINDQUIST, Geraldine Taylor, Ed.D. Utah State University, 1978. 103pp. Major Professor: Dr. Arthur D. Jackson

The purposes of this study were: (1) to determine if Denver Developmental Screening Test scores predicted the Gates-MacGinitie Reading Test scores of primary grade students, (2) to measure the relation of attendance at the Weber District pre-kindergarten program on the reading achievement of primary grade students, and (3) to determine the unique contribution of the variables of age, sex, grade level, and each of the four sections of the screening test to the prediction of reading achievement test scores.

To determine if the Denver Developmental Screening Test predicted the Gates-MacGinitie Reading Test Scores, a Pearson product moment correlation (r) was used. To determine if the DDST predicted better for high achieving or low achieving students the errors of estimate were also analyzed.

Results from the data showed a statistical significance at the .001 level for all three grade levels. The relationship however, was not strong enough to provide practical significance. The correlation was highest for first grade students and lowest for those in the third grade. There was no significant difference in the ability of the DDST to predict reading scores for children in the highest and lowest quartile of the reading tests.

To measure the relation of attendance at the Weber School District pre-kindergarten program to reading achievement, an analysis of covariance was used. Differences in scores on the DDST were statistically controlled. The group means compared were for (1) children attending the preschool, (2) children recommended for but not attending the preschool, and (3) those children who were not recommended.

Results from the data showed a significant difference among the means of the reading achievement test scores for the three groups in grades one and two. A posteriori comparisons of means showed the differences between each of the group means were significant in grade one. The difference was significant in grade two only between those attending preschool and those not recommended. The differences among the means for the three groups was not significant for grade three. The mean scores of the Gates-MacGinitie Reading Tests for those children attending the pre-kindergarten session were lower in all grades for either the children recommended for pre-kindergarten but not attending, or for those children not recommended for the preschool.

To determine the contribution of the variables of sex, age, grade level, on each of the four sections of the DDST to the prediction of reading achievement, a multiple regression analysis was used.

Results from the data showed the Fine Motor-Adaptive section was the best single predictor at all grade levels. The Language section made a significant contribution for the third and first grade. Gross Motor and Personal-Social sections did not make a significant contribution to prediction at any grade level. Sex contributed less as a predictor of reading achievement as children progressed in school. The contribution of age at the time of the DDST to prediction was not statistically significant at any grade level. The combined variables had the highest multiple R for first grade and the value decreased as the length of time increased. However, the differences at the three grade levels was not significant.

AN APPLICATION OF GENERALIZABILITY THEORY TO THE ASSESSMENT OF WRITING ABILITY

Order No. 7907770

LLABRE, Maria Magdalena, Ph.D. The University of Florida, 1978. 91pp. Chairman: William B. Ware

Classical reliability theory, as used in the social sciences, has been restricted by a model which specifies one undifferentiated error component. This restriction has limited the applicability of the model and has obscured its interpretation. Recent advancements in psychometric theory provide more flexible models which permit the investigation of multiple sources of error variation. Under the rubric of generalizability theory, these methods are based on R. A. Fisher's work on the analysis of variance and the factorial experiment.

Generalizability theory is potentially very useful in many areas of research suffering from inconsistency of measurement. In particular, the theory is applicable to the assessment of writing ability from written compositions. However, applied studies in this area are lacking.

The literature on the measurement of writing ability has identified several sources of error affecting the reliability of written compositions. The most common sources of error noted are inconsistency across raters, modes, and occasions. In spite of the recognition of these sources of variation, most researchers who have studied the reliability of written composition have examined the issue only in terms of inter-rater reliability. Implicit in the concept of inter-rater reliability is the assumption that fluctuations among raters is the only source of error in the model. This study incorporated three facets: raters, modes, and occasions, in a split-plot factorial design in order to examine the results obtained by taking into account more than one source of error through the methodology of generalizability theory.

Samples of writing from 104 fourth graders were obtained under selected mode and occasion conditions. Each sample was scored by four trained raters. In the design, the students were considered as nested within a higher classification, the classes. The number of students in each class was not constant. Therefore, this study also extended the principles of generalizability theory to unbalanced designs.

Point estimates of the variance components for all effects in the model were obtained through the MIVQUE method. Negative estimates were replaced by zeros. The relative magnitude of the estimates indicated that students could be differentiated on the basis of their ratings. However, the classes as units could not be distinguished. The estimates also showed that errors resulting from variability in the quality of writing across occasions and modes outweigh those stemming from differences among raters. Furthermore, occasions represented a greater source of error than modes. With training and practice, raters can consistently score the writing samples of students using a general impression method.

Assuming homogeneity of variance, unbiased generalizability coefficients were obtained for seven universes of generalization. These universes represented generalization across one facet, two facets, or all three facets simultaneously. The coefficients indicated that, to obtain acceptable levels of generalizability, at least six samples of writing from each person are necessary.

The standard error of measurement which may be used in constructing confidence intervals around a person's universe score was also examined. The results from this examination paralleled those based on the generalizability coefficients.

A supplementary analysis which allowed a comparison of the estimates obtained through the MIVQUE method to those derived using expected mean squares, resulted in similar values for all estimates in a model without the classes effect. These results were interpreted as lending support to the MIVQUE method.

It was concluded that generalizability theory is very useful for clarifying problems in reliability in the area of writing ability. Furthermore, the theory need not be limited to situations with balanced data. Valid methods of variance component estimation documented in the statistical literature may be used with unbalanced designs.

A STUDY OF THE VALIDITY OF FIVE SHORT FORMS OF THE TEST OF LANGUAGE DEVELOPMENT

Order No. 7900601

MAGEE, Patricia Ann, Ph.D. The University of Texas at Austin, 1978. 118pp. Supervisor: Linda L. Brown

Currently, those who wish to conduct language screen and research have no reliable valid test of oral language that is quickly and easily administered. To ameliorate this problem, the authors of the Test of Language Development (TOLD) have extracted two short forms from it. The short forms (NH-X and NH-Y) were devised through an item validity technique, i.e., each item of the test was correlated with its respective subtest total, and the items evidencing the highest correlations were selected. The purpose of this study was to: (1) determine the equivalence of NH-X and NH-Y with the TOLD and (2) compare these short forms with short forms created using regression analysis (REGRAN), random item selection (RAND), and clinical judgment (CLIN). Two investigations were conducted. The first used 234 protocols selected randomly from the standardization sample of the TOLD. The results of the first investigation were cross validated on a separate sample of 60 children. The data were analyzed by (1) correlating the short form raw scores with the TOLD raw score, (2) prorating the short form raw scores, computing the means and standard deviations of the prorated short forms and the TOLD and testing for significant differences between them and (3) calculating the frequency of Type I and Type II errors for each of the short forms. Analysis of the standardization data show high degrees of relationship for all the short forms at all age levels except for REGRAN at the four year level and CLIN at the four and six year levels. NH-X, NH-Y and REGRAN demonstrated significant mean differences from the TOLD at the younger age levels. The mean scores of REGRAN and CLIN were significantly different from the TOLD mean scores at the eight year level. Analysis of the cross validation sample revealed high correlations for all short forms at all age levels. The mean scores of NH-X did not differ from the TOLD mean score, NH-Y differed significantly at the four year level. CLIN differed at the eight year level. The mean scores for REGRAN and RAND differed significantly from the TOLD mean scores at all age levels. The analysis involving the frequency of Type I and Type II errors indicates that all the short forms evidence more Type I errors than Type II errors. RAND had more errors than the other short forms but this difference was not significant.

PREDICTING READING ACHIEVEMENT IN COLLEGE STUDENTS

Order No. 7908686

MARTIN, Deanna Coleman, Ph.D. University of Missouri - Kansas City, 1978. 98pp.

The relationship between selected academic and personality variables and reading performance scores were investigated. The purpose of this line of inquiry was to set the stage for more clearly defining reading maturity. Subjects consisted of 123 college students, the majority of whom were enrolled in a reading and study skills course. Subjects took the Nelson-Denny Reading Test, Omnibus Personality Inventory, Learning Styles Inventory, Learning Preference Inventory, Manzo Bestiary Inventory, Hidden Figures Test, Difference Inventory, and a cloze passage test. Cumulative grade point averages were also related to reading performance scores. The individual relationships between each instrument in the battery and reading performance were assessed; no two independent variables were combined in the analysis. Stepwise multiple regression was employed as the major statistical tool. Tests for interaction and curvilinearity were also conducted. Correlations significant beyond the .05 level were discussed.

The amount of variance explained by one instrument or a combination of subscales from an instrument ranged from 7% to 46%. The two scales of the Difference Inventory--Feelings and Life Style, and Self and Social Construct--accounted

for 7% of the variance in reading performance scores. Three scales of the Omnibus Personality Inventory, Theoretical Orientation, Personal Orientation, and Response Bias, predicted 11% of the variance. Four of the Learning Preference Inventory scales, Individual Tutoring, Team Teaching, Lecture Learning, and Directed Individual Learning, predicted 24% of the variance. Cumulative grade point average predicted 41% of the variance. Four scales from the Learning Styles Inventory, Combined Academic Expectation Scale, Structure/Organization, Numeric, and People, predicted 42% of the variance. The cloze passage test predicted 46%. Statistically significant relationships between reading performance scores and the Manzo Bestiary Inventory animals, moose, porpoise, penguin, and horse, were obtained. Finally, there was no statistically significant relationship between scores on the Hidden Figures Test and reading performance.

Specific characteristics indicated by the inventories are discussed. Factors related to reading maturity and diagnostic implications are also discussed.

AN INVESTIGATION OF THE CLOZE PROCEDURE AS A MEANS TO FACILITATE GROUPING FOR READING INSTRUCTION Order No. 7817397

NOLAN, Marianne Casey, Ed.D. Temple University, 1978. 110pp.

Purpose and Scope of the Study

The purpose of this investigation was to see if a quick and fairly accurate instructional reading level can be obtained by using a cloze test prepared from graded basal reader material. The problem was undertaken to determine if the prepared cloze test can be used in place of the highly regarded IRI as an efficient method to place groups of classroom children on their instructional level to expedite reading instruction. The study was designed to determine the percentage of children tested who are at the same instructional level using IRI and the Ransom criteria for cloze. Another aspect investigated was that of criteria which might be more applicable to the population than those set by Ransom. This was observed from the viewpoint of grade level as well as instructional level as set by IRI. Finally, cloze scores for passages at the instructional level established by the IRI were compared with those for passages above and below that level to determine whether cloze could differentiate these levels.

Subjects for the study included 90 students randomly selected from grades two, four and six in a Philadelphia public school. Two tests were administered, an Informal Reading Inventory and a cloze test. For both tests at each grade level, different passages within the same story were selected from the mid-section of each reader, one through six. Five questions were formulated in this investigation and analyzed through the data obtained on the cloze. The IRI was used to set the level for cloze testing. Only instructional reading level was sought.

Findings

Calculation of the test data showed for Question 1 that, when reading a cloze passage that was designed at the instructional level as set by IRI, the Ransom criteria (30–49) correctly identified 63.3 percent of the cloze scores.

For Question 2, a range of scores on the cloze was found that appeared to identify more realistically than did the Ransom those children who were at instructional level as set by IRI. This criterial range was 30 to 54. This newly found criterial range correctly identified 80 percent of the cloze scores and was established as the criteria of this study.

For Question 3, an instructional level criterion for each instructional level as set by IRI regardless of actual grade placement was established. This was calculated within the context of the findings of the newly established criteria (30–54), wherein 80 percent of the children tested were found instructional. Instructional level criteria were found at all levels, except Level 1.

The same procedure was used to answer Question 4. Criteria for grade placement regardless of instructional level set by IRI was established at all grade levels tested.

In consideration of the overall accuracy of the cloze in terms of its matching IRI, the 80 percent agreement between levels established by cloze and the IRI, appears encouraging for use of cloze for Levels 2 and above. Consideration must be given to the fact that the newly found criteria expanded that of Ransom, and should be viewed with caution.

Data calculated for Question 5 found variations in terms of accuracy in both levels tested when compared to the IRI instructional level set.

Implications

Implicit in this research is the suggestion for teacher use of cloze for instructional level determination based on the findings of this study.

A replication of this study, and an extension of it using a large population, might provide some confirmation and extension of the test findings.

Also implied is the need for preliminary training to provide familiarity with the cloze test at first instructional level.

A FIVE YEAR FOLLOW–UP STUDY EXAMINING THE VALIDITY OF A READING READINESS DIAGNOSTIC SCREENING BATTERY Order No. 7903879

POPE, Jean Grove, Ed.D. University of Cincinnati, 1977. 106pp. Supervisor: Dr. Barry Lehrer

This study investigated the predictive accuracy of a multiphase screening design for identification of kindergarten children of average and above average ability who might have reading problems. Five hundred and forty-five children were evaluated at the end of kindergarten. The two initial predictive measures were the reading subtest of the Wide Range Achievement Test and Teacher Checklist. From these results the students were tentatively considered High Risk and Low Risk. The Slosson Intelligence Test was administered to the High Risk students. Those scoring below 90 IQ were eliminated from the project. The remaining High Risk students were administered the Meeting Street School Screening Test. Students who scored one standard deviation below the norm or obtained a raw score below 44 on this test were considered High Risk students.

After a five year period, 151 students (46 High Risk; 105 Low Risk) were individually evaluated on the Woodcock Reading Mastery Test. It was hypothesized: 1) there would be no significant difference between the reading scores of the High Risk and Low Risk students after a five year period, 2) the reading subtest of the Wide Range Achievement Test would be a significant predictor for reading after a five year period, and 3) there would be no significant difference between the predictive power of the reading subtest of the Wide Range Achievement Test and the Meeting Street School Screening Test after a five year period.

The results indicate that there is a significant relationship between reading achievement at the end of fifth grade as measured by the Woodcock Reading Mastery Test and reading at the end of kindergarten as measured by the reading subtest of the Wide Range Achievement Test. Furthermore, the results of this five year follow-up investigation suggest that the use of a multiphase screening battery based upon the use of the Wide Range Achievement Reading subtest, Teacher Checklist, Slosson Intelligence Test, and Meeting Street School Screening Test at the kindergarten level was useful in discriminating High and Low risk students in terms of potential reading problems.

THE DEVELOPMENT OF A READING ATTITUDE INSTRUMENT FOR GRADES SEVEN THROUGH TWELVE

Order No. 7903467

RHODY, Regina Tullock, Ed.D. The University of Tennessee, 1978. 69pp. Major Professor: Dr. J. Estill Alexander

The purpose of this study was to develop an instrument to measure the attitudes toward reading of students in grades seven through twelve. A Likert-type scale was selected as the technique that best met the needs for this instrument.

Pilot interviews were conducted to obtain knowledge concerning the reading attitudes of secondary students and to obtain statements suitable for an attitude instrument. Three junior high schools and three high schools from an urban area participated in the pilot interviews. These schools had a normal distribution of ability and socioeconomic levels.

The pilot scale, composed of 33 statements based on the pilot interviews, was administered to one class for each grade level. Revisions were made using an item-analysis procedure. The revised instrument consisted of 25 of the 33 original statements.

The revised instrument was administered to 349 junior and senior high school students in both urban and rural areas. The test-retest method used to check reliability indicated that the instrument was reliable. An acceptable level of validity was established by the item-analysis procedure and by teacher ratings of students with positive and negative attitudes. In addition, the fact that students themselves suggested the item pool is taken by the researcher as a strong indicant of validity.

IDENTIFICATION AND TESTING OF ESSENTIAL SKILLS IN LITERAL READING COMPREHENSION Order No. 7822204

ROBLYER, Margaret Dale, Ph.D. The Florida State University, 1978. 122pp. Major Professors: Nelson J. Towle, Walter Dick

The purposes of this study were to: (a) demonstrate procedures for deriving and testing reading skills, and to (b) provide empirical evidence of the prerequisite relationship of these skills to reading comprehension.

Seven skills hypothesized to be prerequisite to comprehension were derived through an analysis of the Resnick and Beck (1976) information-processing model of reading comprehension. These included: decoding, word recognition (vocabulary), word meaning in context, sentence syntax, sentence meaning in context, function word identification, and anaphoric reference. A performance objective was written for each skill.

Baker's (1977) Decoding Skills Inventory was used as a measure of decoding skill. Instruments were developed to measure performance on each of the other skills at the ninth grade level. A cloze test was also designed to measure reading comprehension at the ninth grade level. The Hidden Patterns Test (HPT) (French, Ekstrom, & Price, 1963) was employed as a measure of non-verbal intellectual processing ability. Investigator-developed instruments and tape-recorded directions were formatively evaluated through an internal review, a one-on-one use, and a small group test.

One hundred seventy-six ninth grade students in English classes from two Leon County, Florida, high schools participated in the study. Students were tested during regular English-class periods on two consecutive days in each school.

Data were analyzed using measures of reliability, multiple regression analyses, and prerequisite analyses suggested by White and Gagné (1978).

All seven skill tests were found to exhibit high reliability (r_{11} = .74 to .92) as measured by a test of internal consistency (KR-20). The cloze test also showed high reliability as measured by a test-retest correlation (r_{12} = .81).

A stepwise regression analysis was done with the seven skill measures and Hidden Patterns Test as independent variables and cloze test performance as the dependent variable. Results of the analysis indicated that the variables were, as a group, good predictors of cloze test performance (R = .86). This analysis indicated that three variables accounted for .72 of the variance of cloze performance and were the only significant contributors: decoding, sentence syntax, and anaphoric reference. Score on the HPT was found to contribute very little. The two lowest reading skill predictors were: word meaning in context and sentence meaning in context.

Prerequisite analyses indicated that all skills were necessary for comprehension as measured by the cloze test. The result of a prerequisite analysis of extreme cloze scores (highest quarter and lowest quarter of scores) indicated that all skills were usually present in good readers and lacking in poor readers. The deficiency in the three significant skill predictors was especially apparent in poor readers.

SOME EFFECTS OF METHODS OF STANDARDIZED READING-ACHIEVEMENT TEST ADMINISTRATION AND SCORE INTERPRETATION ON SENIOR-HIGH STUDENTS' ATTITUDE TOWARD READING Order No. 7824344

SHANNON, Albert Joseph, Ph.D. Marquette University, 1978. 83pp.

The present study was designed to determine the effects of standardized testing on student attitude toward reading. It was hypothesized that the methods surrounding reading-test administration and student achievement on the test would not affect attitude toward reading. An additional hypothesis was formulated indicating that these two independent variables (method of administration and student achievement) interact to affect attitude toward reading.

The Nelson-Denny Reading Test (1973) was used to measure reading achievement. The even-numbered items of the Kennedy and Halinski (1975) "70-Item Attitude Instrument" were used to measure attitude toward reading. The tenth-grade subjects (N = 120) were randomly assigned to one of the four experimental conditions. The Counseled group (n = 30) received pretest counseling and posttest score interpretation surrounding administration of the reading-achievement test. The Score Only group (n = 30) received only their Total score grade equivalent following the administration of the standardized reading test. The No Score group (n = 30) received no score report following testing. The Control group (n = 30) received no standardized reading-achievement test. Following treatment, all groups completed the attitude toward reading scale.

The method of test administration was found to significantly affect student attitude toward reading. Significant differences in attitude were found between the No Score group and all other groups in the design. The No Score method of test administration resulted in the least favorable attitude toward reading. Attitude scores of subjects in the Control group were higher than attitude scores of subjects who were tested but not told their scores. The effects of testing without reporting the scores to the students were sufficiently negative to strongly suggest reassessment of testing done in this manner.

For all subjects who were given the achievement test, the attitude toward reading of the Counseled group was the most positive. Attitude scores for the Counseled group differed significantly from the scores of the Score Only group. In effect, if a test is to be given, and the administrator is conscious of the effects it may have on attitudes toward reading, pretest counseling and posttest score interpretation are recommended.

Of the groups receiving the achievement test, the group with the second most positive attitude toward reading was the Score Only group. Reporting the scores to the students following testing produced a more positive attitude toward reading than not reporting the scores to the students. Reporting only the scores, however, did not produce as positive an attitude toward reading as providing pretest counseling and posttest score interpretations to the students.

Attitude scores of the Counseled group and the Score Only group did not significantly differ from the Control group, which was not given the standardized reading test. Attitude scores of the No Score group, however, did differ significantly from the untested Control group. Data here indicate that the pretest

counseling and posttest score interpretation, or the simple reporting of scores to the students following a standardized test, will not significantly change attitude toward reading. Furthermore, testing students without informing them of their scores will have a significant negative effect on attitude toward reading.

DIFFICULTY OF COMPREHENSION QUESTIONS IN NORM- AND CRITERION-REFERENCED TESTS ADMINISTERED TO FOURTH GRADE PUPILS Order No. 7901688

SMITH, Elizabeth Jane Wooten, Ph.D. University of Georgia, 1978. 289pp. Supervisor: Ira E. Aaron

The purpose of this study was to examine the test items of norm- and criterion-referenced achievement tests of reading comprehension administered at the fourth grade level to determine the frequency of occurrence and differences in difficulty of three levels of comprehension questions, i.e., literal, inference, and evaluation. These broad levels of comprehension questions were categorized further into sublevels or specific comprehension tasks. The frequency of occurrence and differences in difficulty of the comprehension sublevels were also examined.

Data were collected for the criterion-referenced test of reading used in the statewide testing program and for three norm-referenced measures selected by the researcher for 240 subjects. Test content and pupil performance were examined with regard to level of comprehension question and type of test instrument.

Three groups of raters classified test items from the selected instruments into levels and sublevels of comprehension by use of the Taxonomy of Reading Comprehension (Barrett, 1972).

Test passage difficulty of the four instruments was determined by consulting test manuals and by using the Dale-Chall Readability Formula.

Nine hypotheses were investigated. Rater classifications of test items indicated the absence of evaluation questions on three of the four instruments used. Therefore, hypothesis one, dealing with the frequency of occurrence of different levels of comprehension questions was only partially tested. Hypothesis three, dealing with the frequency of occurrence of the sublevels of comprehension, was dealt with similarly due to the absence of several sublevels of comprehension on the tests used. The testing of hypothesis two, which dealt with the percent of correct responses to different levels of comprehension, resulted in mixed findings for the different instruments used. With the exception of the comparison of boys' and girls' scores on inference and evaluation questions on the criterion-referenced test and the comparison of mean scores for the combined sample on literal and inference questions on one norm-referenced test, significant differences in pupil performance were indicated for all levels of questions on all tests.

Significant differences in the percent of correct responses to test questions classified according to sublevels of comprehension were found. A significant relationship was indicated between pupil percent correct scores on different levels of comprehension questions on all tests used. The same finding was indicated for pupil percent correct total scores on the four tests administered. Significant differences in total test scores were found for the combined sample and for the sample categorized by sex, race, and socioeconomic status. Significant differences in pupil scores among items of any one level of comprehension across passages of varying difficulty were found.

Descriptive analyses of grade equivalent scores for each norm-referenced test were conducted. The range of grade equivalent scores for different levels of mastery on the criterion-referenced test was determined.

The results of the study indicate that the Taxonomy of Reading Comprehension (Barrett, 1972) does possess utility as a tool for labeling comprehension tasks. Findings indicate that the suggested ordering of the levels or sublevels of comprehension does not imply an ordering in terms of task difficulty.

Variations in format and content exist among the three norm-referenced tests and one criterion-referenced test administered in the study in terms of test length, passage readability levels, number of items classified into different levels of comprehension, number of comprehension skills tested, and the relative emphasis placed on these skills. Inference level questions and skills appear predominant.

Variability in pupil performance from test to test was indicated for grade equivalent scores on the three norm-referenced tests. For the different levels of mastery on the criterion-referenced test a large range of norm-referenced grade equivalent scores was indicated.

AN INVESTIGATION OF THREE DOMAIN REFERENCED STRATEGIES TO ASSESS SECONDARY STUDENTS' EXPOSITORY WRITING Order No. 7913755

SMITH, Laura Spooner, Ed.D. University of California, Los Angeles, 1979. 247pp. Chairman: Professor Eva Lee Baker

As demand increases for competency based tests of students' basic academic skills, additional requirements for measures of writing proficiency also have surfaced. The requirements call for measures that are technically sound and which serve as meaningful, efficient indicators of clearly defined writing competencies. Additionally, the demand is for measures that carry clear implications for instructional planning.

The present study was undertaken to examine relationships among writing assessment strategies which are potentially responsive to requirements of competency based testing. Three domain referenced strategies to measure secondary students' expository writing were developed. Two of the strategies, direct measures, involved collecting and rating students' writing samples. The distinction between the strategies lay in the form of the essay rating criteria. One form, an Analytic rating scale, yielded separate scores on the following elements: essay focus, development, organization, support, paragraphing, and mechanics. The other form, an Impressionistic rating scale, yielded a single score on the quality of each essay as an example of exposition. The third strategy, an indirect measure, was an objective test of writing-related competencies derived from the Analytic rating scale.

Subjects, 128 eleventh and twelfth grade students in six English classes, were randomly assigned within each class to treatment groups determined by the other in which the measures were administered. Each subject wrote two essays of at least 200 words on topics designed to elicit expository writing and completed the Objective test. Two raters were trained to use the Analytic scale and two to use the Impressionistic scale. The writing samples were scored by both rater pairs, resulting in four scores for each sample.

Findings pertaining to the first area of inquiry, relationships between the two direct measure strategies, revealed that Impressionistic scale, Analytic total scale and Analytic subscale scores were highly related. Two Analytic subscales, Mechanics and Support, were significant predictors to Impressionistic scores. The strong relationships, which suggested that the scales measured comparable elements of writing, may possibly be explained by the high degree of uniformity in the rhetorical structures of the writing samples.

Findings related to the second area of inquiry, relationships between the direct and indirect measures, revealed moderately high correlations between Objective test total scores and total scores of the two rating scales. Two Objective subtests, Paragraphing and Paragraph Analysis, emerged as significant predictors to Impressionistic and Analytic total scores. The Objective subtests, with one exception, were significant predictors to analogous Analytic subscales. The positive relationship between the direct and indirect measures was expected, as the Objective items were designed to assess, at the levels of recognition and discrimination, those categories of skill measured by the rating scales at the level of production.

Data on the subjects' recent instructional history in composition was collected through questionnaires administered to the subjects and their current English teachers. These data revealed minor between class variation in instructional emphases; however, the data did not adequately explain the pattern of class performances across the measures.

THE QUALITY OF STUDENT COMPOSITION AS PREDICTED BY AVERAGE NUMBER OF WORDS PER T-UNIT AND ORGANIZATION SKILLS Order No. 7900117

STOKES, Patricia Welch, Ph.D. Georgia State University – College of Education, 1978. 66pp.

Purpose

The purpose of this research was to determine if any interaction exists between any two of three factors: judged quality of student composition, organizational skills, and length and complexity of sentences.

Methods and Procedures

Ninety students--thirty from grades 8, 10, and 12--wrote four compositions, one each in four modes--descriptive, argumentative, expository, and narrative. The finished papers, in their original form, were rated by three judges for skill in organization and overall quality of writing according to the General Impression Marking system used by the College Entrance Examination Board. A scale of 1 (low) to 4 (high) was used to record the raters' judgments.

The researcher then calculated the average number of words per T-Unit according to the method used by Frank O'Hare and Warren Combs in their sentence combining research.

The data was then subjected to a multiple analysis of variance and covariance via an SPSS system to determine if interaction among the factors existed. When positive results indicated a need to determine more specifically where the interaction lay, a Duncan's Multiple Range Test was employed to gain an indication of the strength of the interaction, if any.

Results

The multiple analysis of variance and covariance showed a significant interaction to exist between organization skills and quality of writing, but no significant interaction existed between either quality of writing and average T-Unit length or organization skills and average T-Unit length. Further information gained was the significant superiority of the narrative mode in the students' writing and the indication that twelfth graders wrote significantly more words per T-Unit than eighth and tenth graders, though eighth and tenth graders wrote approximately the same average number of words per T-Unit.

Conclusions

It was concluded that length and complexity of sentences does not have an effect on the judged quality of student composition, though skill in organization does. Previous research which indicated an increase in sentence length and complexity would yield quality composition is called into question and recommendations for researching methods of teaching organization skills as a tool in teaching composition are included.

ISSUES OF VALIDITY AND RELIABILITY IN THE TESTING OF FRESHMAN COMPOSITION Order No. 7900944

WALLACE, Rose Ann, Ph.D. The University of New Mexico, 1978. 121pp.

An investigation of the construct validity of first language proficiency test scores reveals a substantial common underlying factor in first language tests administered to 138 college freshmen. A principal components analysis suggests that perhaps this underlying factor accounts for more than half of the reliable variance in the test scores. Initial and follow-up tests administered to the freshmen included tests designed to tap skills related to proficiency in reading and writing standard formal English: a writing sample, the Cooperative English Tests (Reading Comprehension and English Expression), and the College English Placement Test. Cloze tests were included to determine whether the cloze procedure might be used as a valid means of assessing proficiency in composition as well as in reading comprehension and second language proficiency. The principal components analysis of initial and follow-up test scores revealed that the cloze test scores loaded significantly on what might be termed a literacy factor; moreover, in the varimax rotated factor solutions, the cloze test scores loaded significantly on a hypothesized writing factor. The cloze procedure may be a valid means of assessing the ability to organize written statements into paragraphs and essays; the loading of the cloze procedure on this factor with essay ratings is an indication of a possible common skill underlying both.

Analyses of co-variance and a critical study of items on the CET, the cloze tests, CEPT provided bases for an evaluation of their predictive and content validity. As a means of assessing predictive validity, analyses of co-variance in scores were computed with initial ability and teaching method as co-variates. Reading sub-tests, judged to be the most valid of the CET and CEPT subtests in terms of construct validity, also appeared to have the greatest predictive validity. The predictive validity of cloze tests was difficult to determine since students did not perform as well over-all on the follow-up test as on the initial test. Those students termed "high ability" achieved greater gains from initial to follow-up testing than those termed "low ability" in composition skills, contrary to the expectation that there would be a ceiling effect.

If assumptions of integrative testing theory, as postulated by second language test theorists, are applicable to first language testing, the CET and CEPT, which were designed according to psychometric test construction theory and notions of content and predictive validity, do not adequately reflect the universe tested (first language reading and writing proficiency and, perhaps, an underlying over-all proficiency in standard English).

In any consideration of the validity of essay ratings as measures of writing ability, the reliability of the ratings must be taken into account since the two are inextricably linked. Consequently, the reliability of essay ratings was investigated under three rating conditions: essays were rated by (1) college professors having no training sessions for the specific rating task, (2) college professors having training sessions using model essays, and (3) college professors having the same training sessions as those in the second group, but rating in pairs. The ratings of trained judges marking in consultation correlated most strongly with other raters as well as with scores on the other measures. In the varimax rotated factor analysis of test scores, the most reliable ratings were included; these loaded on the writing factor with the cloze scores and course grades.

Further research on the reliability of essay ratings might include comparisons of kinds of training sessions useful in increasing rating reliability. In addition, the use of the cloze procedure as an integrative measure for assessing writing ability as well as over-all proficiency in standard English merits investigation. Finally, additional research on the construct validity of measures of first language proficiency may provide a direction for the design, revision, or selection of tests of reading and writing skill in the first language.

A COMPARATIVE STUDY OF READABILITY FORMULAS APPLIED TO COLLEGE TEXTBOOKS Order No. 7912530

WOOD, Westa Winn, Ph.D. Kent State University, 1978. 192pp.
Director: Barbara A. Hutson

It was the purpose of this study to examine the characteristics of four readability formulas by applying them to seven college texts actually used by college freshmen with relatively low reading skills.

The study sought to find answers to the following questions:

1. Do the estimated mean reading levels vary as a function of the specific texts and the different readability formulas applied?

2. Does the variability of estimated reading levels of the various passages differ as a function of the specific formulas applied?

3. Are the estimated reading levels, as determined by various formulas, positively correlated?

Four readability formulas were applied to the same 10 passages in each of seven selected college texts in the social sciences. The four formulas were: (1) McLaughlin's (1969a) SMOG formula; (2) the FORCAST formula (Caylor, Sticht, Fox, & Ford, 1972); (3) the Dale-Chall formula (Dale & Chall, 1948); and (4) the Fry formula, extended version (Fry, 1977).

A two-way analysis of variance design (Formulas (4) x Books (7) with repeated measures on formulas) was used to test hypothesis one. As a follow-up analysis, the Newman-Keuls and Scheffé multiple-comparisons procedures were employed to locate the significant differences in means as indicated by the F ratio obtained. A dependent t-test was used to test hypothesis two, examining the differences in variances. The Pearson product-moment correlation was used to test hypothesis three.

Results

There was a significant difference between means for texts; the mean for one of the books was significantly lower than the means for other books. There was a significant interaction of books with formulas. Most of the interaction involved differences between Book 4 (introductory philosophy) and the other books used.

There were significant differences between the pairs of formulas yielding the highest grade equivalents (i.e., SMOG and Fry) and the pairs of formulas yielding the lowest grade equivalents (i.e., FORCAST and Dale-Chall), but there were no significant differences within the pairs of formulas.

Further analysis suggested that a number of technical words, numbers, proper names, etc., included in texts may result in wide differences in readability estimates for that book by different formulas.

There are indications that some formulas yield more precise estimates, at least in some types of material, than do others. Some ranged from .7 to 2.5 years.

Overall, significant positive correlations were found across all books, but when looked at within books, the correlations were mixed; that is, some correlations were significant and positive; some correlations were moderately high and positive; some correlations were lower positive; and other correlations were low or essentially zero, except for the significant negative correlation between the SMOG and FORCAST formulas for Book 7.

Conclusions

The findings from the study seem to justify the following conclusions.

Estimated mean reading levels of texts may vary, depending on the particular content of the text being assessed, and upon what variables are measured in the formula used and how these variables are measured.

There were indications that some formulas yield more precise estimates than do others. The standard error of the four formulas used ranged from 0.7 to 2.5 years.

While there were significant positive correlations across books for each pair of formulas studied, the size of the correlations was less than one would expect. Within books, correlations ranged from significant positive to significant negative.

Because grade level estimates tend to vary across formulas, the results obtained from any given readability formula should be cautiously interpreted.

Readability formulas may be useful for determining the relative difficulty of materials if kept in proper perspective.

Copies of the dissertations may be obtained by addressing your request to:

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106

or by telephoning (toll-free) 1-800-521-3042